



Heriot-Watt University  
Research Gateway

## Sparse regularization for fiber ODF reconstruction

**Citation for published version:**

Daducci, A, Van De Ville, D, Thiran, J-P & Wiaux, Y 2014, 'Sparse regularization for fiber ODF reconstruction: From the suboptimality of  $\ell_2$  and  $\ell_1$  priors to  $\ell_0$ ', *Medical Image Analysis*, vol. 18, no. 6, pp. 820-833. <https://doi.org/10.1016/j.media.2014.01.011>

**Digital Object Identifier (DOI):**

[10.1016/j.media.2014.01.011](https://doi.org/10.1016/j.media.2014.01.011)

**Link:**

[Link to publication record in Heriot-Watt Research Portal](#)

**Document Version:**

Peer reviewed version

**Published In:**

Medical Image Analysis

**Publisher Rights Statement:**

Copyright © 2014 Elsevier B.V. All rights reserved.

**General rights**

Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [open.access@hw.ac.uk](mailto:open.access@hw.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

# Sparse regularization for fiber ODF reconstruction: from the suboptimality of $\ell_2$ and $\ell_1$ priors to $\ell_0$

A. Daducci<sup>a,\*</sup>, D. Van De Ville<sup>b,d</sup>, J-P. Thiran<sup>a,c</sup>, Y. Wiaux<sup>a,b,d,e</sup>

<sup>a</sup>*Signal Processing Lab (LTS5), École Polytechnique Fédérale de Lausanne, Switzerland*

<sup>b</sup>*Medical Image Processing Lab, École Polytechnique Fédérale de Lausanne, Switzerland*

<sup>c</sup>*University Hospital Center (CHUV) and University of Lausanne (UNIL), Switzerland*

<sup>d</sup>*Department of Radiology and Medical Informatics, University of Geneva, Switzerland*

<sup>e</sup>*Institute of Sensors, Signals & Systems, Heriot-Watt University, Edinburgh, UK*

---

## Abstract

Diffusion MRI is a well established imaging modality providing a powerful way to probe the structure of the white matter non-invasively. Despite its potential, the intrinsic long scan times of these sequences have hampered their use in clinical practice. For this reason, a large variety of methods have been recently proposed to shorten the acquisition times. Among them, spherical deconvolution approaches have gained a lot of interest for their ability to reliably recover the intra-voxel fiber configuration with a relatively small number of data samples. To overcome the intrinsic instabilities of deconvolution, these methods use regularization schemes generally based on the assumption that the fiber orientation distribution (FOD) to be recovered in each voxel is sparse. The well known Constrained Spherical Deconvolution (CSD) approach resorts to Tikhonov regularization, based on an  $\ell_2$ -norm prior, which promotes a weak version of sparsity. Also, in the last few years compressed sensing has been advocated to further accelerate the acquisitions and  $\ell_1$ -norm minimization is generally employed as a means to promote sparsity in the recovered FODs. In this paper, we provide evidence that the use of an  $\ell_1$ -norm prior to regularize this class of problems is somewhat inconsistent with the fact that the fiber compartments all sum up to unity. To overcome this  $\ell_1$  inconsistency while simultaneously exploiting sparsity more optimally than through an  $\ell_2$  prior, we reformulate the reconstruction problem as a constrained formulation between a data term and a sparsity prior consisting in an explicit bound on the  $\ell_0$  norm of the FOD, i.e. on the number of fibers. The method has been tested both on synthetic and real data. Experimental results show that the proposed  $\ell_0$  formulation significantly reduces modeling errors compared to the state-of-the-art  $\ell_2$  and  $\ell_1$  regularization approaches.

*Keywords:* diffusion MRI, HARDI, reconstruction, compressed sensing

---

\*Corresponding author. *Postal address:* EPFL STI IEL LTS5, ELD 232, Station 11, CH-1015 Lausanne (Switzerland). *Phone:* +41 (0) 21 6934622

*Email address:* [alessandro.daducci@epfl.ch](mailto:alessandro.daducci@epfl.ch) (A. Daducci)

---

## 1. Introduction

Fiber-tracking is probably one of the most fascinating applications in diffusion MRI (dMRI), gathering a lot of attention since its introduction because of its ability to reconstruct the main neuronal bundles of the brain from the acquired data. In fact, the random movement of the molecules in the white matter can be exploited for mapping brain connectivity, and structures otherwise invisible with other imaging modalities can be highlighted. The study of this structural connectivity is of major importance in a fundamental neuroscience perspective, for developing our understanding of the brain, but also in a clinical perspective, with particular applications for the study of a wide range of neurological disorders.

The most powerful acquisition modality is diffusion spectrum imaging (DSI) (Wedeen et al., 2005). It relies on cartesian signal sampling and is known to provide good imaging quality, but it is too time-consuming to be of real interest in a clinical perspective. Diffusion tensor imaging (DTI) (Basser et al., 1994) is always preferred instead. DTI is a very fast model-based technique providing valuable diagnostic information but, on the contrary, it is unable to model multiple fiber populations in a voxel. In a global connectivity analysis perspective, this constitutes a key limiting factor. Accelerated acquisitions relying on a smaller number of samples while providing accurate estimations of the intra-voxel fiber configuration thus represent an important challenge.

Recently, an increasing number of high angular resolution diffusion imaging (HARDI) approaches have been proposed for tackling this problem. In particular, spherical deconvolution (SD) based methods formed a very active area in this field (Tournier et al., 2004; Alexander, 2005; Tournier et al., 2007; Dell’acqua et al., 2007). These methods rely on the assumption that the signal attenuation acquired with diffusion MRI can be expressed as the convolution of a given response function with the fiber orientation distribution (FOD). The FOD is a real-valued function on the unit sphere ( $S^2$ ) giving the *orientation* and the *volume fraction* of the fiber populations present in a voxel. The response function, or kernel, describes the dMRI signal attenuation generated by an isolated single fiber population; it can be estimated from the data or represented by means of parametric functions. SD approaches represented a big step in reducing the acquisition time of diffusion MRI, but are known to suffer heavily from noise and intrinsic instabilities in solving the deconvolution problem. For this reason, a regularization scheme is normally employed. A variety of approaches have been proposed, which are generally based on the two assumptions that the FOD is (i) a non-negative function and (ii) sparse, i.e. with only a few nonzero values, either explicitly or implicitly. In fact, at the imaging resolution available nowadays, diffusion MRI is sensitive only to the major fiber bundles and it is commonly accepted that it can reliably *disentangle up to 2-3 different fiber populations* inside a voxel (Jeurissen et al., 2010; Schultz, 2012). Hence, the FOD can reasonably be considered sparse in nature. In particular, the state-of-the-art

Constrained Spherical Deconvolution (CSD) approach of [Tournier et al. \(2007\)](#) resorts to Tikhonov regularization, based on an  $\ell_2$ -norm prior. While its primary purpose is to ensure the positivity of the FOD, it actually also implicitly promotes sparsity, but only a weak version of it.

The recent advent of *compressed sensing* (CS) theory ([Donoho, 2006](#); [Candès et al., 2006](#); [Baraniuk, 2007](#)) provided a mathematical framework for the reconstruction of sparse signals from under-sampled measurements mainly in the context of convex optimization. CS has inspired new advanced approaches in the last few years for solving the reconstruction problem in diffusion MRI and allowed a further dramatic reduction in the number of samples needed to accurately infer the fiber structure in each voxel, by promoting sparsity explicitly. For instance, [Tristán-Vega and Westin \(2011\)](#) and [Michailovich et al. \(2011\)](#) recovered the orientation distribution function (ODF) by using different representations for the response function, while [Merlet et al. \(2011\)](#) and [Rathi et al. \(2011\)](#) focused on the full ensemble average propagator (EAP) of the diffusion process. In this work, however, we focus on spherical deconvolution based-methods and the quantity of interest is the FOD. In general these methods are based on  $\ell_1$  minimization, where the  $\ell_1$  norm is defined as  $\|\mathbf{x}\|_1 = \sum_i^n |\mathbf{x}_i|$  for any vector  $\mathbf{x} \in \mathbb{R}^n$ , and the common goal is to recover the FOD with fewest non-zeros that is compatible with the acquired dMRI data ([Ramirez-Manzanares et al., 2007](#); [Pu et al., 2011](#); [Landman et al., 2012](#); [Mani et al., 2012](#)). However, a minimum  $\ell_1$ -norm prior is inconsistent with the physical constraint that the sum of the volume fractions of the compartments inside a voxel is intrinsically equal to unity.

In this paper, we propose to exploit the versatility of compressed sensing and convex optimization to solve what we understand as *the  $\ell_1$  inconsistency*, while simultaneously exploiting sparsity more optimally than the approaches based on the  $\ell_2$  prior, and improve the quality of FOD reconstruction in the white matter. Our approach is as follows. Strictly speaking, the FOD sparsity is the number of fiber populations, thus identified by the  $\ell_0$  norm of the FOD.  $\ell_0$ -norm problems are generally intractable as they are non convex, which explains the usual convex  $\ell_1$ -norm relaxation in the framework of compressed sensing. To this end, some greedy algorithms have been proposed to approximate the  $\ell_0$  norm through a sequence of incremental approximations of the solution, such as Matching Pursuit ([Mallat and Zhang, 1993](#)) and Orthogonal Matching Pursuit ([Pati et al., 1993](#)). However, the greedy and local nature of these algorithms, i.e. in the sense that compartments are identified sequentially, makes them suboptimal as compared to more robust approaches based on convex optimization, which are global in nature. In particular, a reweighted  $\ell_1$  minimization scheme was developed by [Candès et al. \(2008\)](#) in order to approach  $\ell_0$  minimization by a sequence of convex weighted- $\ell_1$  problems. We thus solve the  $\ell_0$  minimization problem by making use of a reweighting scheme and evaluate the effectiveness of the proposed formulation in comparison with state-of-the-art approaches based on either  $\ell_2$  or  $\ell_1$  priors. We report results on both synthetic and real data.



## 2. Materials and methods

### 2.1. Intra-voxel structure recovery via spherical deconvolution

As shown by [Jian and Vemuri \(2007\)](#), spherical deconvolution methods can be cast into the following computational framework:

$$S(b, \hat{\mathbf{q}})/S_0 = \int R_{\hat{\mathbf{q}}}(\hat{\mathbf{p}}) f(\hat{\mathbf{p}}) d\Omega(\hat{\mathbf{p}}), \quad (1)$$

where  $f$  is the FOD to be estimated,  $R_{\hat{\mathbf{q}}}$  the response function rotated in direction  $\hat{\mathbf{q}} \in \mathbb{S}^2$  and the integration is performed over the unit sphere with  $\hat{\mathbf{p}} = (\phi, \theta) \in \mathbb{S}^2$  and  $d\Omega = \sin \phi d\phi d\theta$ .  $S(b, \hat{\mathbf{q}})$  represents the dMRI signal measured on the q-space shell acquired with b-value  $b$  in direction  $\hat{\mathbf{q}} \in \mathbb{S}^2$ , while  $S_0$  is the signal acquired without diffusion weighting. The FOD  $f$  is normally expressed as a linear combination of basis functions, e.g. spherical harmonics, as  $f(\hat{\mathbf{p}}) = \sum_j w_j f_j(\hat{\mathbf{p}})$ . The measurement process can thus be expressed in terms of the general formulation:

$$\mathbf{y} = \Phi \mathbf{x} + \eta, \quad (2)$$

where  $\mathbf{x} \in \mathbb{R}_+^n$  are the coefficients of the FOD,  $\mathbf{y} \in \mathbb{R}^m$  is the vector with the dMRI signal measured in the voxel with  $y_i = S(b, \hat{\mathbf{q}}_i)/S_0$  for  $i \in \{1, \dots, m\}$ ,  $\eta$  represents the acquisition noise and  $\Phi = \{\phi_{ij}\} \in \mathbb{R}^{m \times n}$  is the observation matrix modeling explicitly the convolution operator with the response function  $R$ , with  $\phi_{ij} = \int R_{\hat{\mathbf{q}}_i}(\hat{\mathbf{p}}) f_j(\hat{\mathbf{p}}) d\Omega(\hat{\mathbf{p}})$ . Several choices for the convolution kernels and basis functions exist in the literature; more details will be provided on the specific  $\Phi$  used with each algorithm considered in this work.

### 2.2. $\ell_2$ prior

In the original formulation of [Tournier et al. \(2004\)](#), the FOD  $\mathbf{x}$  and the measurements  $\mathbf{y}$  were expressed by means of spherical harmonics (SH), and the deconvolution problem was solved by a simple matrix inversion. To reduce noise artifacts, a low-pass filter was applied for attenuating the high harmonic frequencies. The method was improved in [Tournier et al. \(2007\)](#) by reformulating the problem as an iterative procedure where, at each iteration, the current solution  $\mathbf{x}^{(t)}$  is used to drive to zero the negative amplitudes of the FOD at the next iteration with a Tikhonov regularization ([Tikhonov and Arsenin, 1977](#)):

$$\mathbf{x}^{(t+1)} = \underset{\mathbf{x}}{\operatorname{argmin}} \|\Phi \mathbf{x} - \mathbf{y}\|_2^2 + \lambda^2 \|L^{(t)} \mathbf{x}\|_2^2, \quad (3)$$

where  $\|\cdot\|_p$  are the usual  $\ell_p$  norms in  $\mathbb{R}^n$ , the free parameter  $\lambda$  controls the degree of regularization and  $L^{(t)}$  can be understood as a simple binary mask preserving only the directions of negative or small values of  $\mathbf{x}^{(t)}$ . The  $\ell_2$ -norm regularization term therefore tends to send these values to zero, as probably spurious, hence favoring large positive values. Interestingly, beyond the claimed purpose of enforcing positivity, the operator  $L$  thus also implicitly promotes a weak version of sparsity. However, this  $\ell_2$  prior does not explicitly guarantee either positivity or sparsity in the recovered FOD. In [Alexander \(2005\)](#) a maximum-entropy

regularization was proposed to recover the FOD as the function that exhibits the minimum information content. The method showed higher robustness to noise than previous approaches, but was limited by the very high computational cost and did not promote sparsity. Other regularization schemes have been proposed in the literature, but FOD sparsity has never been addressed with a rigorous mathematical formulation.

### 2.3. $\ell_1$ prior

Compressed sensing provides a powerful mathematical framework for the reconstruction of sparse signals from a low number of data (Donoho, 2006; Candès et al., 2006), mainly in the context of convex optimization. According to this theory, it is possible to recover a signal from fewer samples than the number required by the Nyquist sampling theorem, provided that the signal is sparse in some *sparsity basis*  $\Psi$ . Let  $\mathbf{x} \in \mathbb{R}^n$  be the signal to be recovered from the  $m \ll n$  linear measurements  $\mathbf{y} = \Phi \mathbf{x} \in \mathbb{R}^m$  and  $\boldsymbol{\alpha} \in \mathbb{R}^n$  a sparse representation of  $\mathbf{x}$  through  $\Psi \in \mathbb{R}^{n \times n}$ . If the observations  $\mathbf{y}$  are corrupted by noise and  $\Phi$  obeys some randomness and incoherence conditions, then the signal  $\mathbf{x} = \Psi \boldsymbol{\alpha}$  can be recovered by solving the convex  $\ell_1$  optimization problem:

$$\underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \|\boldsymbol{\alpha}\|_1 \quad \text{subject to} \quad \|\Phi \Psi \boldsymbol{\alpha} - \mathbf{y}\|_2 \leq \epsilon, \quad (4)$$

where  $\epsilon$  is a bound on the noise level. Assuming Gaussian noise, the square  $\ell_2$  norm of the residual represents the log-likelihood of the data and follows a  $\chi^2$  distribution. For a sufficiently large number of measurements, this distribution is extremely concentrated around its mean value. This fact is related to the well-known phenomenon of concentration of measure in statistics. Consequently,  $\epsilon$  can be precisely defined by the mean of the  $\chi^2$ .

In the context of FOD reconstructions, the sparsity basis  $\Psi$  boils down to the identity matrix, thus  $\mathbf{x} = \boldsymbol{\alpha}$ . In Ramirez-Manzanares et al. (2007) and Jian and Vemuri (2007) the sensing basis  $\Phi$ , also called *dictionary*, is generated by applying a set of rotations to a given Gaussian kernel (i.e. diffusion tensor) and the sparsest coefficients  $\mathbf{x}$  of this linear combination best matching the measurements  $\mathbf{y}$  are recovered by solving the following constrained minimization problem:

$$\underset{\mathbf{x} \geq 0}{\operatorname{argmin}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\Phi \mathbf{x} - \mathbf{y}\|_2 \leq \epsilon, \quad (5)$$

where the positivity constraint on the FOD values was directly embedded in the formulation of the convex problem. For  $\text{SNR} > 2$  the noise in the magnitude dMRI images can be assumed Gaussian-distributed<sup>1</sup> (Gudbjartsson and Patz, 1995). However, a statistical estimation of  $\epsilon$  is not reliable, precisely because the number of measurements is very small and the  $\chi^2$  is not really concentrated

---

<sup>1</sup>Since dMRI data is commonly normalized by the baseline  $S_0$ , the  $S_0$  image must be accurately estimated in order to keep the same noise statistics of the non-normalized signal. This is normally the case, though, as multiple  $S_0$  volumes are commonly acquired in practice.

around its mean value. Thus,  $\epsilon$  becomes an arbitrary parameter of the algorithm. At very low SNR, one can also extrapolate the choice of the  $\ell_2$  norm as a simple penalization term independent of statistical considerations.

The reconstruction problem can also be re-formulated as a regularized (as opposed to constrained)  $\ell_1$  minimization as in [Landman et al. \(2012\)](#) and [Pu et al. \(2011\)](#):

$$\underset{\mathbf{x} \geq 0}{\operatorname{argmin}} \|\Phi \mathbf{x} - \mathbf{y}\|_2^2 + \beta \|\mathbf{x}\|_1, \quad (6)$$

where the free parameter  $\beta$  controls the trade-off between the data and the sparsity constraints. In general,  $\beta$  depends on the acquisition scheme and the noise level and it must be empirically optimized. Following the general CS approach, problems (5) and (6) consider an  $\ell_1$ -norm prior on the FOD  $\mathbf{x}$ . However, in the dMRI context, a minimum  $\ell_1$ -norm prior is inconsistent with the physical constraint that the sum of the volume fractions of the compartments inside a voxel is intrinsically equal to unity, i.e.  $\|\mathbf{x}\|_1 \equiv \sum_i \mathbf{x}_i = 1$ . For this reason, we reckon that also these  $\ell_1$ -based formulations are intrinsically suboptimal. Fig. 1 illustrates this inconsistency by reporting the  $\ell_1$  norm of reconstructed FODs as a function of the amplitude of measurement noise.

Our main goal in this work is to demonstrate the suboptimality of the approaches based on  $\ell_2$  and  $\ell_1$  priors and to suggest a new formulation, based on an  $\ell_0$  prior, adequately characterizing the actual sparsity lying in the FOD.

#### 2.4. $\ell_0$ prior

In the aim of adequately characterizing the FOD sparsity, we re-formulate the reconstruction problem as a constrained  $\ell_0$  minimization problem:

$$\underset{\mathbf{x} \geq 0}{\operatorname{argmin}} \|\Phi \mathbf{x} - \mathbf{y}\|_2^2 \text{ subject to } \|\mathbf{x}\|_0 \leq k, \quad (7)$$

where  $\|\cdot\|_0$  explicitly counts the number of nonzero coefficients and  $k$  represents an upper bound on the expected number of fiber populations in a voxel.

As already stated, the  $\ell_0$  problems as such are intractable. The reweighting scheme proposed by [Candès et al. \(2008\)](#) proceeds by sequentially solving weighted  $\ell_1$  problems of the form (7), where the  $\ell_0$  norm is substituted by a weighted  $\ell_1$  norm defined as  $\|\mathbf{w}\boldsymbol{\alpha}\|_1 = \sum_i \mathbf{w}_i |\boldsymbol{\alpha}_i|$ , for positive weights  $\mathbf{w}_i$  and where  $i$  indexes vector components. At each iteration, the weights are set as the inverse of the values of the solution of the previous problem, i.e.  $\mathbf{w}_i^{(t)} \approx 1/\mathbf{x}_i^{(t-1)}$ . At convergence, this set of weights makes the weighted  $\ell_1$  norm independent of the precise value of the nonzero components, thus mimicking the  $\ell_0$  norm while preserving the tractability of the problem with convex optimization tools. Of course, it is not possible to have infinite weights for null coefficients; so a stability parameter  $\tau$  must be added to the coefficients in the selection of the weights.

The *main steps of the reweighted scheme* are reported in the algorithm 1; in the remaining of the manuscript we will refer to it as L2L0, as it is based on a  $\ell_0$  prior. We empirically set  $\tau = 10^{-3}$  and the procedure was stopped if  $\frac{\|\mathbf{x}^{(t)} - \mathbf{x}^{(t-1)}\|_1}{\|\mathbf{x}^{(t-1)}\|_1} < 10^{-3}$  between two successive iterations or after 20 iterations.

---

**Algorithm 1** Reweighted  $\ell_1$  minimization for intra-voxel structure recovery

---

**Input:** Diffusion MRI signal  $\mathbf{y} \in \mathbb{R}^m$  and sensing basis  $\Phi \in \mathbb{R}^{m \times n}$

**Output:** FOD  $\mathbf{x} \in \mathbb{R}^n$

Set the initial status:

$t \leftarrow 0$  and  $\mathbf{w}_i^{(0)} \leftarrow 1, i = 1, \dots, n$  (the symbol  $\leftarrow$  denotes assignment)

**repeat**

Solve the problem:

$$\mathbf{x}^{(t)} \leftarrow \underset{\mathbf{x} \geq 0}{\operatorname{argmin}} \|\Phi \mathbf{x} - \mathbf{y}\|_2^2 \text{ subject to } \|\mathbf{w}^{(t)} \mathbf{x}\|_1 \leq k$$

Update the weights:

$$\mathbf{w}_i^{(t+1)} = \frac{1}{|\mathbf{x}_i^{(t)}| + \tau}$$

$t \leftarrow t + 1$

**until** stopping criterion is satisfied

$\mathbf{x} \leftarrow \mathbf{x}^{(t-1)}$

---

At the first iteration the weighted  $\ell_1$  norm is the standard  $\ell_1$  norm given  $\mathbf{w} = 1$ , and therefore the constraint  $\|\mathbf{w}^{(0)} \mathbf{x}\|_1 \leq k$  is a weak bound on the sum of the fiber compartments and does not constitute a limitation in the procedure.

The proposed  $\ell_0$  approach thus strongly promotes sparsity (by opposition with the  $\ell_2$  approach) and circumvents the  $\ell_1$  inconsistency. It is noteworthy that our formulation at least partially addresses the problem of arbitrary parameters such as  $\epsilon$  in (5) and  $\beta$  in (6), or  $\lambda$  in (3). Our parameter  $k$  indeed explicitly identifies an upper bound on the number of fibers. As discussed before and largely assumed in the literature, we can expect to have at maximum 2-3 fiber compartments in each voxel. The algorithm was found to be quite robust to the choice of  $k$ , and differences were not observed for values up to  $k = 5$ .

Finally, an explicit constraint  $\sum_i \mathbf{x}_i = 1$  might have been added, as it represents the physical property that the volume fractions must sum up to unity. For the sake of simplicity, in this work this constraint was not included (as it is always the case), assuming it is carried over by the data and well-designed bases as pointed out by Ramirez-Manzanares et al. (2007). In sections 3.1.1 and 3.2.3 we will provide evidence that actually this physical constraint is not met when using  $\ell_2$  or  $\ell_1$  priors, whereas it is correctly satisfied with our proposed  $\ell_0$  formulation. This might have severe consequences on the reconstruction quality.

### 2.5. Comparison framework

We compared our  $\ell_0$  approach based on problem (7) against state-of-the-art  $\ell_2$  and  $\ell_1$  approaches respectively based on problems (3) and (6), and referred to as L2L2 and L2L1. To run L2L2 reconstructions we made use of the original `mrtrix` implementation of Tournier et al. (2012), setting the optimal parameters as suggested by the software itself. To solve the L2L1 and L2L0 problems we used the SPArse Modeling Software (SPAMS)<sup>2</sup>, an open-source toolbox written in C++ for solving various sparse recovery problems. SPAMS contains a very

---

<sup>2</sup><http://spams-devel.gforge.inria.fr>

fast implementation of the LARS algorithm (Efron et al., 2004) for solving the LASSO problem and its variants as the L2L1 problem in equation (6) and the weighted  $\ell_1$  minimizations required for our L2L0 approach in equation (7). Numerical simulations on synthetic data were performed to quantitatively assess the performance of L2L2, L2L1 and L2L0 under controlled conditions. The effectiveness of the three priors was also assessed in case of real human brain data.

## 2.6. Numerical simulations

Independent voxels with two fiber populations crossing at specific angles ( $30^\circ - 90^\circ$  range) and with equal volume fractions were synthetically generated. The signal  $S$  corresponding to each voxel configuration was simulated by using the exact expression given in Soderman and Jonsson (1995) for the dMRI signal attenuation from particles diffusing in a restricted cylindrical geometry of radius  $\rho$  and length  $L$  with free diffusion coefficient  $D_0$ . The following parameters were used (Özarslan et al., 2006; Jian and Vemuri, 2007):  $L = 5$  mm,  $\rho = 5$   $\mu$ m,  $D_0 = 2.02 \times 10^{-3}$  mm<sup>2</sup>/s,  $\Delta = 20.8$  ms,  $\delta = 2.4$  ms. The signal  $S$  was contaminated with *Rician noise* (Gudbjartsson and Patz, 1995) as follows:

$$S_{\text{noisy}} = \sqrt{(S + \xi_1)^2 + (\xi_2)^2}, \quad (8)$$

where  $\xi_1, \xi_2 \sim \mathcal{N}(0, \sigma^2)$  and  $\sigma = S_0/\text{SNR}$  corresponds to a given signal-to-noise ratio on the  $S_0$  image. We assumed  $S_0 = 1$  without loss of generality. Because of this assumption, we have implicitly considered a constant echo-time for acquisitions with different b-values, thus ignoring the fact that higher b-values normally require longer echo-times and therefore the images have a lower signal-to-noise ratio. The study of the impact of the echo-time on different regularization priors is beyond the scope of our investigation.

For each voxel configuration, the signal was simulated at different b-values,  $b \in \{500, 1000, \dots, 4000\}$  s/mm<sup>2</sup>, and seven q-space sampling schemes were tested, respectively with 6, 10, 15, 20, 25, 30 and 50 samples equally distributed on half the unit sphere using electrostatic repulsion (Jones et al., 1999) assuming antipodal symmetry in diffusion signal. Six different noise levels were considered,  $\text{SNR} = 5, 10, \dots, 30$ . For every SNR, 100 repetitions of the same voxel were generated using different realizations of the noise. In our experiments, the actual signal-to-noise ratio in the simulated signal was always in a range where the Gaussian assumption on the noise holds. In the extreme setting with a  $\text{SNR} = 5$  on the  $S_0$  and  $b = 4000$  s/mm<sup>2</sup> the actual signal-to-noise ratio in the diffusion weighted signal was about 1.4.

## 2.7. Evaluation criteria

As one of the aims of this work is to improve SD reconstructions, we adopted standard metrics widely used in the literature (Ramirez-Manzanares et al., 2008; Landman et al., 2012; Michailovich et al., 2011) to assess the quality of the reconstructions with respect to number and orientation of the fiber populations:

- *Probability of false fiber detection.* This metric quantifies the correct assessment of the real number  $M$  of populations inside a voxel:

$$P_d = \frac{|M - \tilde{M}|}{M} \cdot 100\%, \quad (9)$$

where  $\tilde{M}$  is the estimated number of compartments. As  $P_d$  does not distinguish between missed fibers and extra compartments found by the reconstruction, we also make use of the following two quantities where needed,  $n^-$  and  $n^+$ , explicitly counting the number of *under-* and *over-*estimated compartments, respectively.

- *Angular error.* This metric quantifies the angular accuracy in the estimation of the directions of the fiber populations in a voxel:

$$\epsilon_\theta = \frac{180}{\pi} \arccos(|\mathbf{d} \cdot \tilde{\mathbf{d}}|), \quad (10)$$

where  $\mathbf{d}$  is a true direction and  $\tilde{\mathbf{d}}$  is its closest estimate. The final value is an average over all fiber compartments by first matching the estimated directions to the ground-truth without using twice the same direction.

Peaks detection was performed using a local maxima search algorithm on the recovered FOD, considering a neighborhood of orientations within a cone of  $15^\circ$  around every direction. For this reason, evaluation metrics are not sensitive for small crossing angles and results are reported in a conservative range  $30^\circ$ – $90^\circ$ . To filter out spurious peaks, values smaller than 10% of the largest peak were discarded; in the case of L2L2 we had to increase this threshold to 20%, as suggested in [Tournier et al. \(2007\)](#), in order to compare with the other methods.

## 2.8. Real data

The human brain data have been acquired from 3 young healthy volunteers on a 3T Magnetom Trio system (Siemens, Germany) equipped with a 32-channel head coil using standard protocols routinely used in clinical practice. Each dataset corresponds to a distinct subject. Two DTI scans (referred in the following as **dti**<sub>30</sub> and **dti**<sub>20</sub>) were acquired at  $b = 1000$  s/mm<sup>2</sup> using 30 and 20 diffusion gradient directions, respectively, uniformly distributed on half the unit sphere using electrostatic repulsion ([Jones et al., 1999](#)). Other acquisition parameters were as follows: TR/TE = 7000/82 ms and spatial resolution =  $2.5 \times 2.5 \times 2.5$  mm for dataset **dti**<sub>30</sub>, while TR/TE = 6000/99 ms and spatial resolution =  $2.2 \times 2.2 \times 3$  mm for dataset **dti**<sub>20</sub>. One HARDI dataset (referred as **hardi**<sub>256</sub>) was acquired at  $b = 3000$  s/mm<sup>2</sup> using 256 directions uniformly distributed on half the unit sphere ([Jones et al., 1999](#)), TR/TE = 7000/108 ms and spatial resolution =  $2.5 \times 2.5 \times 2.5$  mm. To study the robustness of the three algorithms to different under-sampling rates, the **hardi**<sub>256</sub> dataset has been retrospectively under-sampled and two additional datasets (**hardi**<sub>50</sub> and **hardi**<sub>20</sub>) have been created, consisting of only 50 and 20 diffusion directions, respectively. These subsets of directions were randomly selected in order to be as

much equally distributed on half the unit sphere as possible. The actual SNR in the  $b = 0$  images, computed as the ratio of the mean value in a region-of-interest placed in the white matter and the standard deviation of the noise estimated in the background, was about 60 in `dti30`, 30 in `dti20` and 30 in `hardi256`.

### 2.9. Implementation details

In all our experiments, the *response function* was estimated from the data following the procedure described in [Tournier et al. \(2007\)](#). A different response function was estimated for every combination of experimental conditions (number of samples, b-value, SNR), which was then used consistently in the three reconstruction methods. Specifically, the 300 voxels with the highest fractional anisotropy were selected as expected to contain only one fiber population, and a tensor was fitted from the dMRI signal in each. In the case of numerical simulations, an additional set of data containing 300 voxels with a single fiber compartment was generated for this scope. The estimated coefficients were then averaged to provide a robust estimation of the signal profile for the response function. As we used the tool `estimate_response` of `mrtrix` for these operations, the estimated kernel was already suitable to be fed into the L2L2 algorithm. Note that the fiber directions rely on a maxima identification from the SH coefficients, which can take any continuous position on the sphere. Conversely, in the case of both L2L1 and L2L0, the estimated kernel was used to create the dictionary  $\Phi$  by rotating it along 200 orientations uniformly distributed on half the unit sphere. Because of this discretization, the resulting grid resolution is about  $10^\circ$  and thus the intrinsic average error when measuring the angular accuracy is about  $5^\circ$ . In other words, the precision of both L2L1 and L2L0 is limited by the resolution of the grid used to construct the dictionary. For this reason differences between methods below this threshold will be considered not significant. Note that, to improve the precision it would be sufficient to increase the number of directions of the discretization which, however, would have serious consequences on the efficiency and stability of the minimization algorithm. Interestingly, recent works of [Tang et al. \(2012\)](#) and [Candés and Fernandez-Granda \(2012\)](#) explored a novel theory of CS with continuous dictionaries, in the context of which FOD peaks could be thought to be located with infinite precision. This topic will be the subject of future research. Finally, in order to model adequately any partial/full contamination with cerebrospinal fluid (CSF) that may occur in real data, an *additional isotropic compartment* has been considered by adding a column to  $\Phi$ . This compartment was estimated by fitting an isotropic tensor in voxels within the lateral ventricles.

The free parameter controlling the degree of regularization had to be estimated for both L2L2 and L2L1 algorithms. For the former we used the default values suggested in the original implementation available in the `mrtrix` software. For the latter, the regularization parameter  $\beta$  was empirically estimated following the guidelines of [Landman et al. \(2012\)](#), in order to place the method in its best conditions. In numerical simulations, we created an additional training dataset for every combination of experimental conditions (number of samples, b-value, SNR) and 50 reconstructions were performed varying the parameter

$\beta$  from  $10^{-4} \beta^*$  to  $\beta^*$ , with  $\beta^* = \|2\Phi^T \mathbf{y}\|_\infty$  computed independently in each voxel. The value providing the best reconstructions (according to the above metrics) was then used to run L2L1 on the actual data used for the final comparison. We did not observe any improvement in the reconstructions outside this range. In real data, we tested different values for  $\beta$  but, as the ground-truth is unknown, the optimal value was chosen on the basis of a qualitative inspection of the reconstructions considering their shape, spatial coherence and adherence to anatomy. Nonetheless, we found that the algorithm was quite robust to the choice of  $\beta$  and the value providing visually the best results was always very close to  $\beta = 0.1 \cdot \beta^*$ , as suggested in the same work. Therefore this value was used in all real data experiments. This stability might be probably due to the adaptive strategy of estimating  $\beta^*$  in each voxel from the signal  $\mathbf{y}$ . As already emphasized, L2L0 *does not require any free parameter* to be tuned. In fact, in numerical simulations  $k$  can be fixed in all iterations to 3 while we can safely assume  $k = 5$  in real data, hence larger than the 2-3 fibers normally assumed.

### 3. Results and discussion

#### 3.1. Numerical simulations

We quantitatively compared the three approaches on synthetic data with the aim of assessing the impact on the reconstructions of each regularization scheme (i.e.  $\ell_2$ ,  $\ell_1$  and  $\ell_0$  priors) under controlled conditions. In particular, the quality of the reconstructions was evaluated using the metrics introduced above and selectively varying (i) the number of samples and (ii) the b-value of the acquisition scheme, (iii) the noise level and (iv) the crossing angle between the fiber compartments. Results are reported independently for each experimental condition.

##### 3.1.1. Volume fractions and $\ell_1$ norm

As previously stated, the physical constraint that the volume fractions sum to unity is normally omitted in every problem formulation, as it is expected to be carried over by the data and properly designed bases (Ramirez-Manzanares et al., 2007). In Fig. 1 we explicitly tested whether this property is actually satisfied by the algorithms considered in this work. A more detailed analysis of the performance of each prior is performed in the following sections.

The figure reports the average value for the sum of the volume fractions of the reconstructed FODs (i.e.  $\|\mathbf{x}\|_1$ ), as a function of the noise level, for two acquisition schemes with 30 samples at  $b = 1000$  s/mm<sup>2</sup> and  $b = 3000$  s/mm<sup>2</sup>, respectively. The impact on the reconstructions is shown by means of the normalized mean-squared error  $\text{NMSE} = \|\mathbf{y} - \hat{\mathbf{y}}\|_2^2 / \|\mathbf{y}\|_2^2$  (Michailovich et al., 2011) between the measured signal  $\mathbf{y}$  and its estimate  $\hat{\mathbf{y}}$ . The image clearly demonstrates that both L2L2 and L2L1 reconstructions do not fulfill the  $\sum_i \mathbf{x}_i = 1$  physical constraint, as the sum of the recovered volume fractions always tends to be over-estimated by L2L2 and under-estimated by L2L1. This is a clear effect



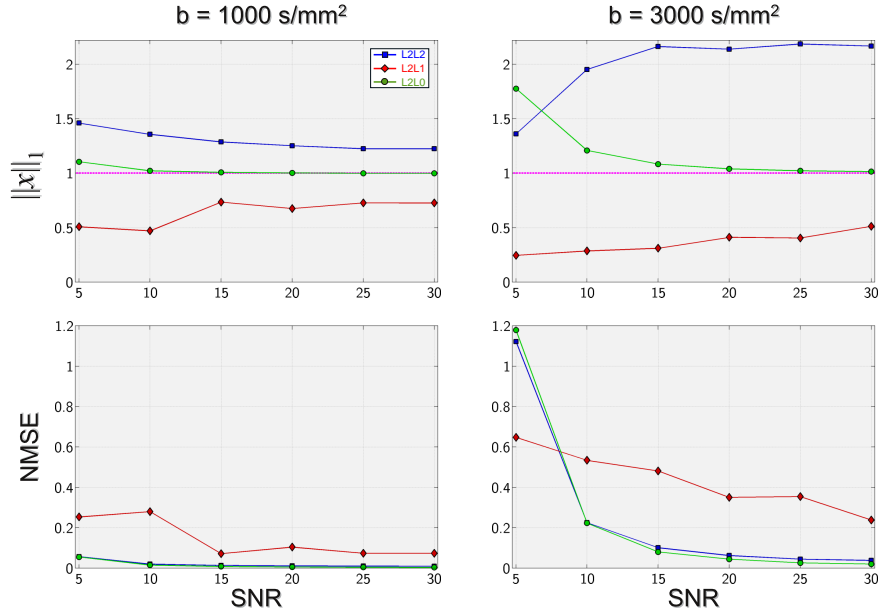


Figure 1: *Sum of volume fractions and impact on the reconstructions.* Top plots report the  $\|\mathbf{x}\|_1$  of the FODs reconstructed by L2L2 (blue boxes), L2L1 (red diamonds) and L2L0 (green circles), while the NMSE of the recovered signal is shown at the bottom. The reference value  $\|\mathbf{x}\|_1 = 1$  is plotted in magenta. Results are reported as a function of the SNR in 2 experimental settings with 30 samples:  $b = 1000 \text{ s/mm}^2$  (left) and  $b = 3000 \text{ s/mm}^2$  (right).

of the weakness of the sparsity constraint in the L2L2 approach and of the inconsistency of the  $\ell_1$  prior in L2L1. On the contrary our L2L0 approach appears to correctly satisfy the constraint, with deviations from unity only with very high noise levels (SNR  $\approx 5$ ). With high quality data this over/under-estimation behavior is fairly mild (at SNR = 30,  $\|\mathbf{x}\|_1 \approx 0.7$  for L2L1 and  $\|\mathbf{x}\|_1 \approx 1.2$  for L2L2), but it progressively intensifies as the noise level increases. The trend is even amplified with high b-value data, in which case the  $\|\mathbf{x}\|_1$  can be as high as  $\approx 2.1$  for L2L2 and as low as  $\approx 0.25$  for L2L1.

Despite showing quite different behaviors with respect to the  $\|\mathbf{x}\|_1$ , L2L2 and L2L0 exhibit very similar NMSE values. On the contrary, L2L1 shows significantly higher reconstruction errors than both L2L2 and L2L0, pointing to the aforementioned  $\ell_1$  *inconsistency*. Debiasing methods (Zou, 2006) have been proposed with the aim to correct the magnitude of the recovered coefficients and mitigate this effect. Nonetheless, a critical step for applying these techniques consists in the proper identification of the support of the solution, otherwise this procedure can lead to really bad results. As we will show in the next sections, this is the case in this work, as the three methods differ significantly in their ability to estimate the number of fiber populations. As the very same data and reconstruction basis have been used for all the methods, we can conclude that any deviation from the unit sum has to be attributed to the different reg-

ularization employed in each algorithm. In the following *we will investigate the consequences on the reconstructions* of using different regularization schemes.

### 3.1.2. Comparison as a function of the number of samples

Fig. 2 reports the performance of the three reconstruction methods as the number of samples changes. We considered seven acquisition schemes from 6 to 50 samples and results are reported for a standard scenario, specifically a shell at  $b = 2000$  s/mm<sup>2</sup> with a SNR = 25. The dependence on the b-value and the robustness to noise will be investigated in detail in the following sections. The quality metrics are reported here as the average value computed over all simulated crossing angles (30°–90°).

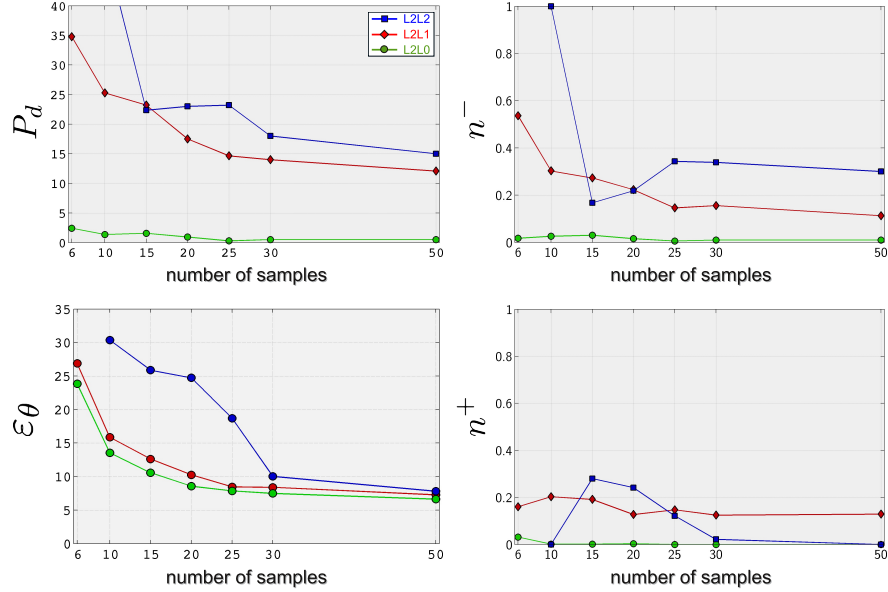


Figure 2: *Quantitative comparison as a function of the number of samples.* The values of the four quality metrics are reported for L2L2 (blue boxes), L2L1 (red diamonds) and L2L0 (green circles) as the number of samples changes. Values shown here correspond to an experimental setting with  $b = 2000$  s/mm<sup>2</sup> and SNR = 25.

Looking at the plots the benefits of using an  $\ell_0$  prior are clear: L2L0 always outperforms both L2L2 and L2L1 in identifying the correct number of fiber populations ( $P_d$ ) and results are consistent for all number of samples considered. The main benefit of L2L0 seems to be the drastically decreased number of missed fibers (smaller  $n^-$ ), even though also the number of over-estimated compartments ( $n^+$ ) is significantly reduced. Concerning the angular accuracy of the recovered fiber populations ( $\epsilon_\theta$ ), reconstructions with L2L0 always resulted in smaller errors as compared to both L2L2 and L2L1. Although the difference with respect to L2L1 is not significant as always within the intrinsic grid precision, both methods showed a substantial improvement over L2L2,

which appeared to suffer from a sudden and significant deterioration of the reconstructions ( $\approx 10^\circ\text{--}15^\circ$ ) for less than 30 samples. This can be explained with the SH representation used internally by L2L2. In fact, even though the FOD is a function on the sphere containing high-resolution features by definition, a maximum SH order  $l_{max} = 4$  (or less) can be used for acquisitions with less than 30 samples, hence drastically reducing the intrinsic angular resolution of the recovered FOD. At least 30 to 60 samples are normally advised for using L2L2, so in our experiments we have actually tested L2L2 beyond its applicability range. On the contrary, L2L1 and L2L0 do not make use of SH and the reconstruction quality degrades more smoothly with the under-sampling rate of the dMRI data. In the following we will focus on two acquisition schemes to further analyze the performance of three methods: (i) in a normal setting with 30 samples and (ii) in a regime of high under-sampling with only 15 samples.

### 3.1.3. Comparison as a function of the crossing angle

In Fig. 3 the performances of L2L2, L2L1 and L2L0 are plotted in detail as a function of the crossing angle between the fiber populations. Results are shown for two acquisitions with 30 and 15 samples, both simulated at  $b = 2000 \text{ s/mm}^2$  and  $\text{SNR} = 25$ .

With 30 samples, the major source of errors for both L2L2 and L2L1 is represented by *under-estimation* ( $n^-$ ), although spurious orientations are not negligible ( $n^+ \approx 0.2$ ). In particular, both methods start to severely miss fibers for crossing angles below  $60^\circ$ , where they tend to recover a single peak lying between the two real fiber directions. In these situations, the maximum angular error for the sole estimated peak is generally upper bounded by half the angle separating the two fibers; for this reason the overall  $\epsilon_\theta$  performances of L2L2 and L2L1 do not differ significantly from L2L0 despite the drastic improvement in terms of  $P_d$ ,  $n^-$  and  $n^+$ . On the other hand, in an under-sampling scenario with 15 samples L2L2 and L2L1 exhibit much higher  $P_d$  values and a stronger tendency to *over-estimate* compartments, usually in completely arbitrary orientations not even close to the true fiber directions. The overall improvement in the angular accuracy of L2L0 is more evident, with an average enhancement up to  $5^\circ$  with respect to L2L1, whereas L2L2 exhibits a severe drop of the performance mainly due to modeling limitations, as previously pointed out.

These differences can have dramatic consequences for fiber-tracking applications. In fact, tractography algorithms are particularly prone to these estimation inaccuracies, i.e. number and orientation of fiber populations, because the propagation of these (perhaps locally small) errors can lead to completely wrong final trajectories. For instance, a missed compartment might stop prematurely a trajectory, while a spurious peak might lead to create an anatomically incorrect fiber tract. Hence, the ability to accurately recover the intra-voxel fiber geometry is of utmost importance.

### 3.1.4. Comparison as a function of the b-value

So far L2L2, L2L1 and L2L0 have been compared for given acquisition schemes at a fixed  $b = 2000 \text{ s/mm}^2$ . Fig. 4 reports the quality of the reconstructions

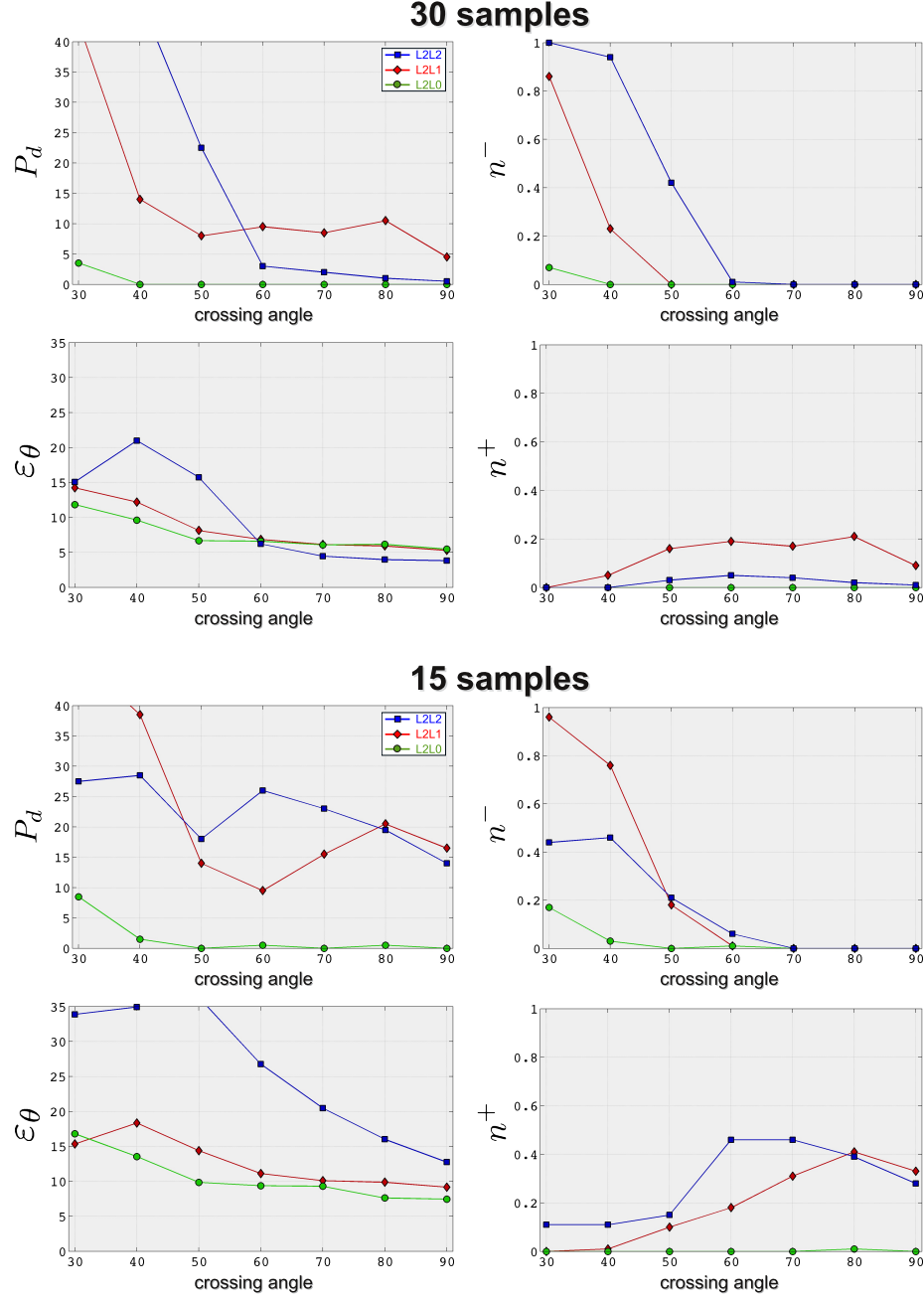


Figure 3: *Quantitative comparison as a function of the crossing angle.* The performances of the three reconstruction methods are detailed separately for each crossing angle used in the simulations. Results are reported for 30 and 15 samples, using the same experimental configuration of Fig. 2, i.e.  $b = 2000 \text{ s/mm}^2$  and  $\text{SNR} = 25$ .

with the three approaches as a function of the  $b$ -value. The results are shown for 30 and 15 samples with a SNR = 25.

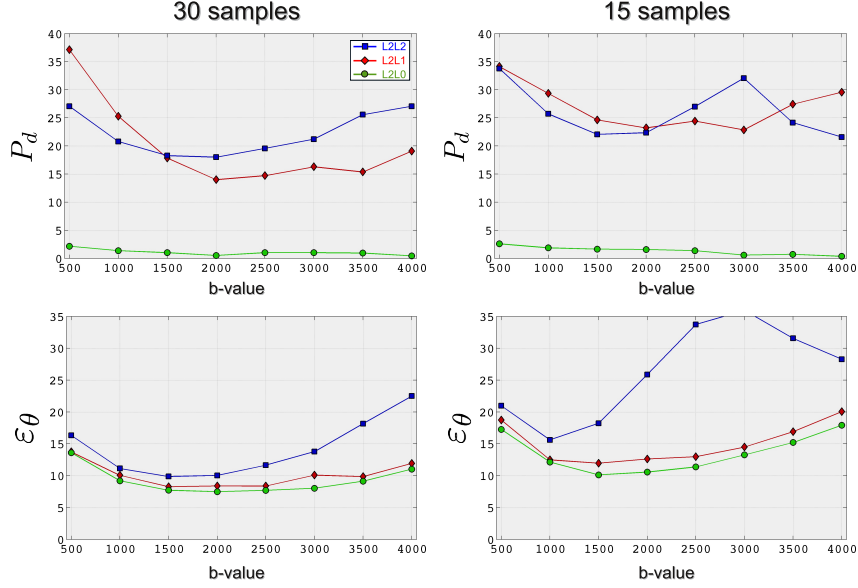


Figure 4: *Quantitative comparison as a function of the  $b$ -value.* The dependence of the reconstruction quality on the  $b$ -values used in the acquisition is reported here for 30 and 15 samples with a SNR = 25.

L2L2 tends to miss compartments for low  $b$ -values and over-estimate them at higher  $b$  ( $n^+$  and  $n^-$  are not shown here for brevity). This is even more apparent when decreasing the number of samples in the acquisition to 15, where L2L2 estimates a lot of spurious peaks at high  $b$ -values (high  $n^+$ ) and thus the angular accuracy of the estimated fiber directions drops considerably. Interestingly, L2L1 shows the opposite behavior, under-estimating at high  $b$  and over-estimating at low  $b$ , although at a smaller rate thus preventing the performance to degrade significantly. Again, in comparison, L2L0 shows a very stable estimation of the number of fibers. Concerning the angular accuracy, all methods showed a minimum for  $\epsilon_\theta$  corresponding to  $b \approx 1500 - 2500$  s/mm<sup>2</sup>, representing a sort of trade-off between the loss in angular resolution happening at small  $b$ -values and the stronger noise influence at higher  $b$ . In fact, as in this work we report the noise level as the SNR of the  $S_0$  dataset, images at high  $b$ -values will have lower actual signal-to-noise ratio, and thus the noise effects will be inherently stronger. Overall, L2L0 always results in smaller angular errors than the other two methods. The improvement with respect to L2L1 is not significant, while the difference with L2L2 is much more pronounced (up to 20°) especially as the  $b$ -value increases.

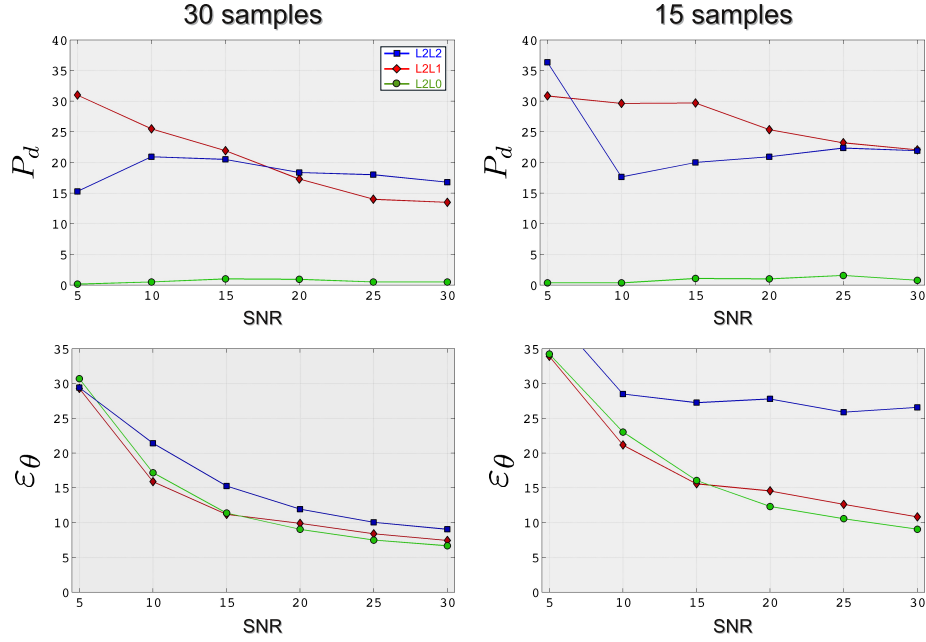


Figure 5: *Quantitative comparison as a function of the SNR.* The robustness to noise of the three reconstruction methods is shown for 30 and 15 samples at  $b = 2000$  s/mm<sup>2</sup>. Reported values for the SNR correspond to the signal-to-noise ratio of the  $S_0$  dataset.

### 3.1.5. Comparison as a function of the SNR

Finally, Fig. 5 compares the robustness to noise of the three methods. Six noise levels have been considered, with the SNR of the  $S_0$  dataset varying from 5 to 30. The comparison is reported for 30 and 15 samples at  $b = 2000$  s/mm<sup>2</sup>. The results show that L2L0 clearly outclasses the other two methods concerning the estimation of the number of compartments ( $P_d$ ) and results are consistent as the SNR changes, both with 30 and 15 samples. In terms of angular accuracy, L2L0 and L2L1 have very similar  $\epsilon_\theta$  performances, almost indistinguishable from one another. On the contrary, L2L2 systematically obtains significantly higher  $\epsilon_\theta$  values at all considered SNRs (up to 6° with 30 samples). In a high under-sampling regime (right plots), the angular accuracy drastically degrades in the case of L2L2 and it appears almost independent of the noise level. This is again consistent with the limitations of the SH representation for acquisitions with very few samples.

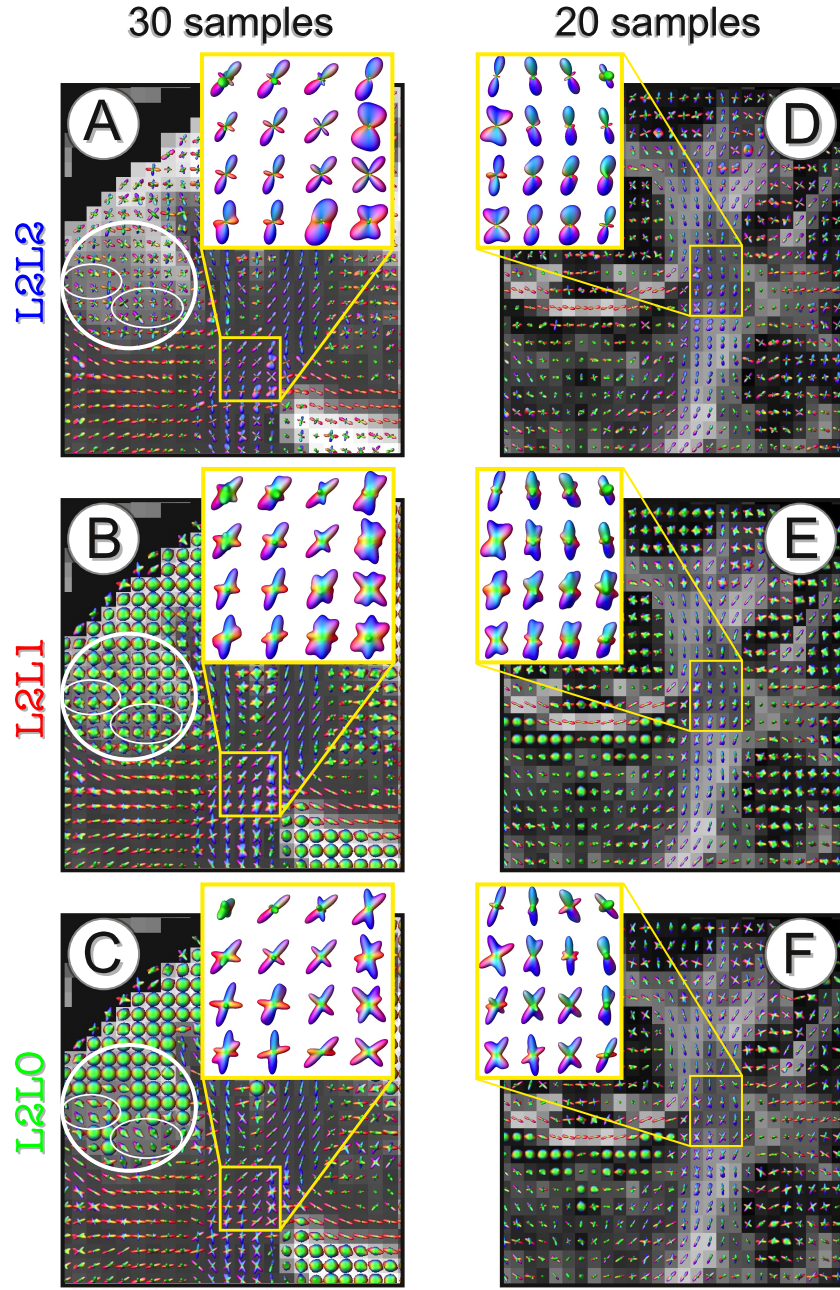


Figure 6: *Qualitative comparison on DTI human brain data.* Reconstructions of the FODs in the corona radiata region are shown for: L2L2 (A and D), L2L1 (B and E) and L2L0 (C and F). FODs in subplots A–C correspond to dMRI images acquired using 30 samples, superimposed on the ADC map, while D–F are relative to the acquisition with 20 samples, superimposed on the FA map. All images have been acquired at  $b = 1000 \text{ s/mm}^2$ .



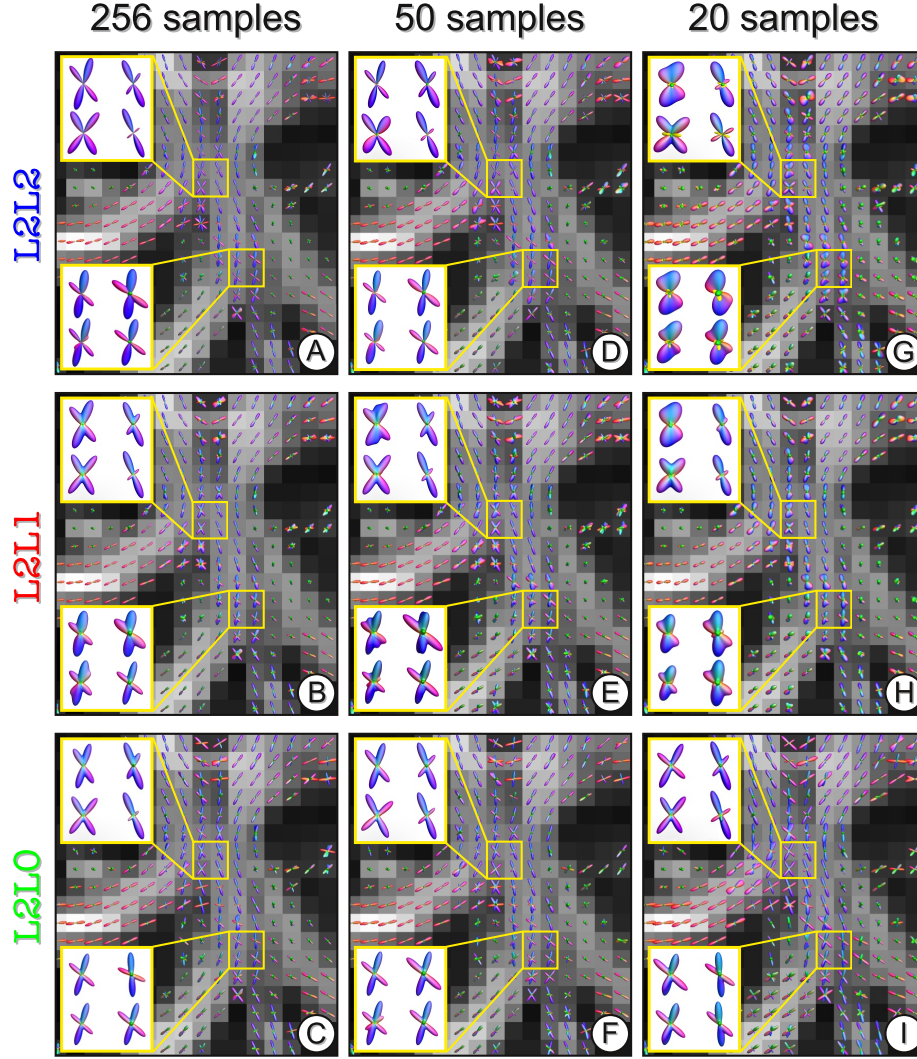


Figure 7: *Qualitative comparison on HARDI human brain data.* Reconstructions of the FODs in the corona radiata region are shown for: L2L2 (A, D and G), L2L1 (B, E and H) and L2L0 (C, F and I). Subplots A–C correspond to the fully-sampled dataset `hardi256` (256 samples), D–F to the dataset `hardi50` (50 samples) while G–I are relative to `hardi20` (20 samples). Images have been acquired at  $b = 3000 \text{ s/mm}^2$ .



### 3.2. Real data

#### 3.2.1. Qualitative evaluation on DTI data

Fig. 6 compares the reconstructions<sup>3</sup> obtained with the three regularization schemes in the case of real data acquired with a typical DTI protocol. Subplots A, B and C correspond to the `dti30` dataset acquired using 30 samples. Even though the acquisition scheme used for this dataset is not the setting where our numerical simulations highlighted the most substantial differences between the three methods, important conclusions can be drawn in favor of L2L0. Looking at the regions in the white circles, the ability of both L2L1 and L2L0 to properly model the isotropic compartment in voxels with full or partial contamination with CSF is clearly visible. On the contrary, as L2L2 does not explicitly model any CSF compartment, it appears unable to adequately characterize the signal in these cases, but it rather approximates any isotropic contribution with a set of random and incoherent fiber compartments. Besides, comparing B and C we can observe that L2L0 successfully differentiates gray matter (light gray regions) from CSF voxels with pure isotropic and fast diffusion (very bright areas), whereas L2L1 appears unable to distinguish them.

The yellow frames highlight the corona radiata, a well-known region in the white matter containing crossing fibers. As expected from our simulations at this still relatively high number of samples, differences are not obvious between the three methods. However, we observe that L2L0 clearly results in sharper and more defined profiles than L2L1, whereas the improvements with respect to L2L2 are confined only to few voxels. The not so good performance of L2L1 might be related to the value chosen for  $\beta$ . In contrast, no free parameter has to be empirically optimized in our approach. When decreasing the acquisition samples to 20 (subplots D, E and F corresponding to `dti20` dataset), fiber directions are definitely much better resolved with L2L0 than with both L2L2 and L2L1. In fact L2L2 clearly breaks, missing many fiber compartments probably due to the aforementioned limitations of the SH representation. The same happens to L2L1, whose reconstructions appear very blurred and noisy.

#### 3.2.2. Qualitative evaluation on HARDI data

The comparison with high b-value data is reported in Fig. 7. The figure shows also the robustness to different under-sampling rates of each scheme. Subplots A, B and C correspond to the fully-sampled dataset `hardi256`. In this situation, no evident differences between the three approaches can be observed as they perform essentially the same. With moderate under-sampled data (subplots D, E and F corresponding to `hardi50`) both L2L2 and L2L0 do not show any significant difference in the quality of the reconstructions, so far exposing neat and sharp profiles. On the other hand, the FODs reconstructed by L2L1 show some signs of progressive degradation, appearing a little more blurred as

---

<sup>3</sup>The images have been created using the tool `mrview` of `mrtrix`. As a consequence, the FODs from L2L1 and L2L0 had to be converted to SH, and this operation caused some blur in the sparse reconstructions of these two methods.

compared to those reconstructed from fully-sampled data (compare subplots E and B). The situation changes drastically with highly under-sampled data, as easily noticeable by comparing the subplots G, H and I, which correspond to the reconstructions performed with only 8% of the original data. In fact, while L2L0 does not show yet any significant degradation of the FODs, both L2L2 and L2L1 clearly do not provide as sharp and accurate reconstructions as in the case of fully-sampled data (compare G to A and H to B). In addition, in the case of L2L2 we can observe a higher incidence of negative peaks (identified in the plots by small yellow spikes), a clear sign of augmented modeling errors.

### 3.2.3. Quantitative comparison: volume fractions and $\ell_1$ norm

In Fig. 8 we tested whether the physical constraint of unit sum is satisfied also in case of real data. The images confirm the observations previously made with synthetic data (cf. Fig 1). In fact, the sum of the recovered volume fractions tends to be over-estimated by L2L2 (subplots A and B) and under-estimated by L2L1 (subplots C and D), whereas L2L0 reconstructions (subplots E and F) appear to meet the property of unit sum as expected. All methods coherently show a mild over-estimation in the corpus callosum, compatible with the highly-packed axonal structure in this region. Finally, L2L2 seems to suffer from over-estimation more with fully- than with under-sampled data, which might be related to the SH order employed for different number of samples.

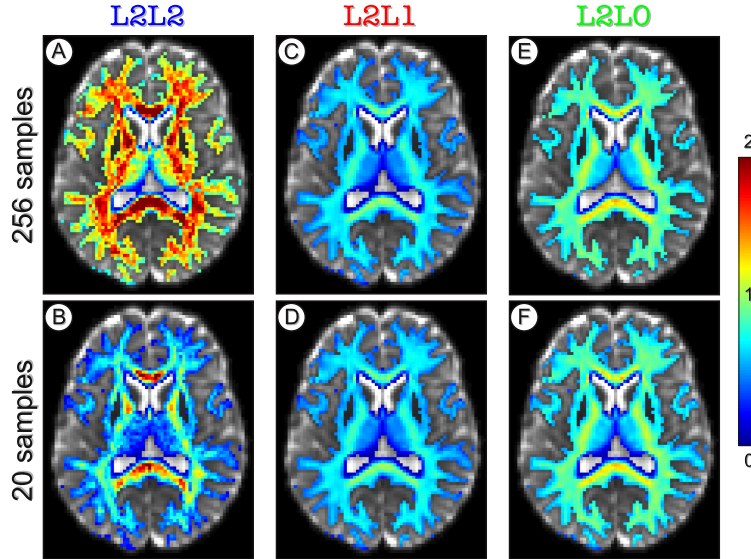


Figure 8: *Sum of volume fractions in real data.* The sum of the volume fractions of the FODs reconstructed with L2L2 (A and B), L2L1 (C and D) and L2L0 (E and F) is reported in a representative slice of the high b-value data. The top row corresponds to the fully-sampled dataset (`hardi256`) while the bottom to the under-sampled one with 20 samples (`hardi20`).

### 3.2.4. Quantitative comparison: fully- vs under-sampled data

We compared the reconstructions obtained from under-sampled data (i.e. **hardi**<sub>50</sub> and **hardi**<sub>20</sub>) to those with fully-sampled data (i.e. **hardi**<sub>256</sub>), considering this latter as ground-truth, as done by [Yeh and Tseng \(2013\)](#). In agreement with the results from numerical simulations, no significant difference was found between the three approaches in terms of angular accuracy. The average error using 50 samples was  $10.9^\circ \pm 9.9^\circ$  (mean  $\pm$  standard deviation) for L2L2,  $8.8^\circ \pm 8.1^\circ$  for L2L1 and  $10.0^\circ \pm 11.3^\circ$  for L2L0. The reconstructions using 20 samples clearly showed higher angular errors. The differences between L2L1 and L2L0 were below the resolution of the sphere discretization used in this study:  $11.6^\circ \pm 9.1^\circ$  and  $12.6^\circ \pm 12.4^\circ$  respectively. L2L2 revealed significantly higher  $\epsilon_\theta$  values:  $17.2^\circ \pm 12.8^\circ$ . On the other hand, results definitely confirmed the superior performance of L2L0 in terms of  $P_d$  that was previously observed in synthetic experiments. With 50 samples L2L0 had an average  $P_d = 4.0\% \pm 13.9\%$  as opposed to sensibly higher values for L2L2 and L2L1, respectively  $17.8\% \pm 32.6\%$  and  $17.3\% \pm 24.3\%$ . For 20 samples, the performance of L2L2 and L2L1 visibly deteriorated,  $42.1\% \pm 43.6\%$  for the former and  $21.3\% \pm 27.7\%$  for the latter. L2L0 reconstructions appeared very stable with an average  $P_d = 5.5\% \pm 15.8\%$ . These enlightening results are illustrated in Fig. 9.

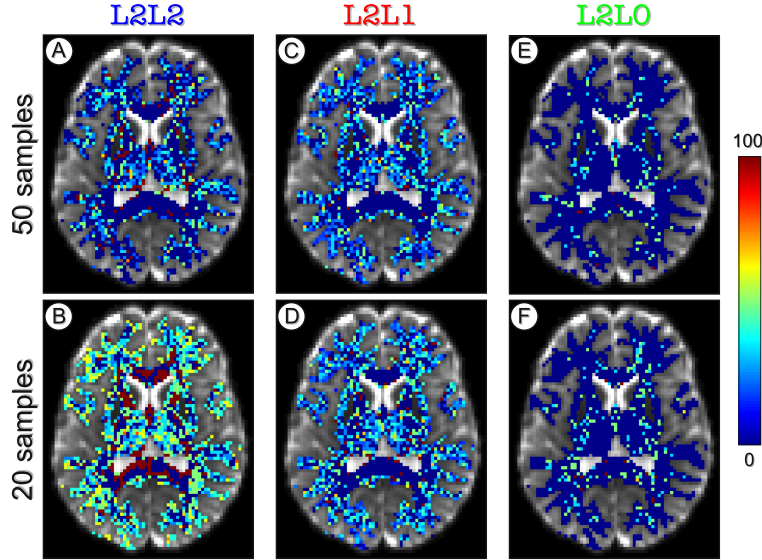


Figure 9: *Comparison between fully- and under-sampled real data.* The performance of L2L2 (A and B), L2L1 (C and D) and L2L0 (E and F) was quantified by considering **hardi**<sub>256</sub> as ground-truth and computing  $P_d$  for the reconstructions on under-sampled data. The top row corresponds to the dataset with 50 samples (**hardi**<sub>50</sub>) and the bottom to 20 (**hardi**<sub>20</sub>).

### 3.3. Limitations and future work

Our proposed formulation represents an extension of classical spherical deconvolution and sparse reconstruction methods (Tournier et al., 2007; Landman et al., 2012) and, as such, it also inherits all the intrinsic limitations and shortcomings of this class of techniques. Like all its predecessors, in fact, our method is based on the assumption that the response function can be estimated from the data and especially that it is adequate for characterizing the diffusion process in all the voxels of the brain. Moreover, the validity of these approaches has yet to be properly assessed with more critical intra-voxel configurations (Sotiropoulos et al., 2012) or pathological brain conditions. Yet, as these methods are currently widely used in this field, we have shown in this work that by *expressing adequately the regularization prior used for promoting sparsity* the quality of the reconstructions can significantly be improved, with no additional cost.

Some of the aforementioned limitations might be addressed by enhancing the estimation of the dictionary accounting for more complex configurations, such as using different response functions for different brain regions and/or pathological tissues and including specific kernels which explicitly model fiber fanning/bending. In addition, even though we focused here on single voxel experiments, future work will be devoted to study the applicability and the effectiveness of our approach in more sophisticated frameworks exploiting the spatial coherence of the data. Finally, future research will investigate the use of the recently proposed continuous CS theory (Tang et al., 2012; Candès and Fernandez-Granda, 2012) with the aim of further improving the accuracy of the reconstructions and reducing the acquisition time.

## 4. Conclusion

In this paper we focused on spherical deconvolution methods currently used in diffusion MRI for recovering the FOD and estimating the intra-voxel configuration in white matter. In particular, we investigated the effectiveness of state-of-the-art regularization schemes based on  $\ell_2$  and  $\ell_1$  priors and provided evidence that these formulations are intrinsically suboptimal: the former because it does not explicitly promote sparsity in the FOD, the latter because it is inconsistent with the fact that the fiber compartments must sum up to unity. We proposed a formulation that rather places a strict bound on the number of expected fibers in the voxel through a bound on the  $\ell_0$  norm of the FOD, relying on a reweighted  $\ell_1$  scheme. We compared our L2L0 approach with the state-of-the-art L2L2 and L2L1 methods, both on synthetic and real human brain data. Results showed that our proposed formulation significantly improves single-voxel FOD reconstructions, with no additional overheads. This evolution is most remarkable in a high q-space under-sampling regime, thus driving the acquisition cost of HARDI closer to DTI.

## Acknowledgments

This work was supported by the Center for Biomedical Imaging (CIBM) of the Geneva and Lausanne Universities, EPFL, the Leenaards and Louis-Jeantet foundations, the EPFL-Merck Serono Alliance award and the Swiss National Science Foundation (SNSF) under grant PP00P2-123438.

## References

- Alexander, D.C., 2005. Maximum entropy spherical deconvolution for diffusion MRI, in: *Information Processing in Medical Imaging (IPMI)*, pp. 76–87.
- Baraniuk, R., 2007. Compressive sensing. *IEEE Signal Processing Magazine* 24, 118–121.
- Basser, P., Mattiello, J., Le Bihan, D., 1994. MR diffusion tensor spectroscopy and imaging. *Biophysical journal* 66(1), 259–267.
- Candès, E., Fernandez-Granda, C., 2012. Towards a mathematical theory of super-resolution. *ArXiv e-prints* [1203.5871](#).
- Candès, E., Romberg, J., Tao, T., 2006. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory* 52(2), 489–509.
- Candès, E., Wakin, M., Boyd, S., 2008. Enhancing sparsity by reweighted  $\ell_1$  minimization. *Journal of Fourier Analysis and Applications* 14(5), 877–905.
- Dell’acqua, F., Rizzo, G., Scifo, P., Clarke, R., Scotti, G., Fazio, F., 2007. A model-based deconvolution approach to solve fiber crossing in diffusion-weighted MR imaging. *IEEE Transactions on Biomedical Engineering* 54, 462–472.
- Donoho, D., 2006. Compressed sensing. *IEEE Transactions on Information Theory* 52(4), 1289–1306.
- Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., 2004. Least angle regression. *The Annals of statistics* 32, 407–499.
- Gudbjartsson, H., Patz, S., 1995. The Rician distribution of noisy MRI data. *Magnetic Resonance in Medicine* 34(6), 910–914.
- Jeurissen, B., Leemans, A., Jones, D., Tournier, J., Sijbers, J., 2010. Estimating the number of fiber orientations in diffusion MRI voxels : a constrained spherical deconvolution study, in: *International Society for Magnetic Resonance in Medicine (ISMRM)*, p. 573.
- Jian, B., Vemuri, B.C., 2007. A unified computational framework for deconvolution to reconstruct multiple fibers from diffusion weighted MRI. *IEEE Transactions on Medical Imaging* 26, 1464–1471.

- Jones, D.K., Horsfield, M.A., Simmons, A., 1999. Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance imaging. *Magnetic Resonance in Medicine* 42, 515–525.
- Landman, B.A., Bogovic, J.A., Wan, H., Elshahaby, F.E.Z., Bazin, P.L., Prince, J.L., 2012. Resolution of crossing fibers with constrained compressed sensing using diffusion tensor MRI. *NeuroImage* 59(3), 2175–2186.
- Mallat, S., Zhang, Z., 1993. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing* 41, 3397–415.
- Mani, M., Jacob, M., Guidon, A., Liu, C., Song, A., Magnotta, V., Zhong, J., 2012. Acceleration of high angular and spatial resolution diffusion imaging using compressed sensing, in: *IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 326–329.
- Merlet, S., Cheng, J., Ghosh, A., Deriche, R., 2011. Spherical Polar Fourier EAP and ODF reconstruction via compressed sensing in diffusion MRI, in: *IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 365–371.
- Michailovich, O., Rath, Y., Dolui, S., 2011. Spatially regularized compressed sensing for high angular resolution diffusion imaging. *IEEE Transactions on Medical Imaging* 30(5), 1100–1115.
- Özarslan, E., Shepherd, T., Vemuri, B., Blackband, S., Mareci, T., 2006. Resolution of complex tissue microarchitecture using the diffusion orientation transform (DOT). *NeuroImage* 31, 1086–1103.
- Pati, Y., Rezaeiifar, R., Krishnaprasad, P.S., 1993. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition, in: *27th Asilomar Conference on Signals, Systems and Computers*, pp. 40–4.
- Pu, L., Trouard, T.P., Ryan, L., Huang, C., Altbach, M.I., Bilgin, A., 2011. Model-based compressive diffusion tensor imaging, in: *IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 254–257.
- Ramirez-Manzanares, A., Cook, P.A., Gee, J.C., 2008. A comparison of methods for recovering intra-voxel white matter fiber architecture from clinical diffusion imaging scans, in: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 305–312.
- Ramirez-Manzanares, A., Rivera, M., Vemuri, B., Carney, P., Mareci, T., 2007. Diffusion basis functions decomposition for estimating white matter intravoxel fiber geometry. *IEEE Transactions on Medical Imaging* 26, 1091–1102.
- Rath, Y., Michailovich, O., Setsompop, K., Bouix, S., Shenton, M.E., Westin, C.F., 2011. Sparse multi-shell diffusion imaging, in: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 58–65.

- Schultz, T., 2012. Learning a reliable estimate of the number of fiber directions in diffusion MRI, in: Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 493–500.
- Soderman, O., Jonsson, B., 1995. Restricted diffusion in cylindrical geometry. *Journal of Magnetic Resonance, Series A* 117, 94–97.
- Sotiropoulos, S.N., Behrens, T.E., Jbabdi, S., 2012. Ball and rackets: Inferring fiber fanning from diffusion-weighted MRI. *NeuroImage* 60, 1412–1425.
- Tang, G., Narayan Bhaskar, B., Shah, P., Recht, B., 2012. Compressed Sensing off the Grid. *ArXiv e-prints* [1207.6053](#).
- Tikhonov, A.N., Arsenin, V.Y., 1977. Solutions of ill-posed problems. New York: Winston & Sons.
- Tournier, J.D., Calamante, F., Connelly, A., 2007. Robust determination of the fibre orientation distribution in diffusion MRI: Non-negativity constrained super-resolved spherical deconvolution. *NeuroImage* 35, 1459–1472.
- Tournier, J.D., Calamante, F., Connelly, A., 2012. MRtrix: Diffusion tractography in crossing fiber regions. *International Journal of Imaging Systems and Technology* 22, 53–66.
- Tournier, J.D., Calamante, F., Gadian, D., Connelly, A., 2004. Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *NeuroImage* 23, 1176–1185.
- Tristán-Vega, A., Westin, C.F., 2011. Probabilistic ODF estimation from reduced HARDI data with sparse regularization, in: Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 182–190.
- Wedeen, V., Hagmann, P., Tseng, W.Y., Reese, T., Weisskoff, R., 2005. Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. *Magnetic Resonance in Medicine* 54, 1377–1386.
- Yeh, F.C., Tseng, W.Y., 2013. Sparse solution of fiber orientation distribution function by diffusion decomposition. *PLoS One* 8.
- Zou, H., 2006. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association* 101, 1418–1429.